

Optimal Polyspectra Estimation

J.R. Fergusson and E.P.S. Shellard

*Centre for Theoretical Cosmology,
Department of Applied Mathematics and Theoretical Physics,
University of Cambridge, Wilberforce Road,
Cambridge CB3 0WA, United Kingdom*

(Dated: January 20, 2013)

We show that full inverse covariance weighting can be naturally incorporated into modal estimation methods making them optimal for the CMB power spectrum, bispectrum and trispectrum, as well as in other 3D applications. Modal estimation methods are highly efficient requiring inversion of only an $n_{\max} \times n_{\max}$ covariance matrix, where n_{\max} is the limited number of modes needed to describe both the theoretical model and the noise which affects it (in contrast to the full $l_{\max}^2 \times l_{\max}^2$ inverse covariance weighting). For an isotropic modal bispectrum estimator implemented at a WMAP resolution, we demonstrate how the modal inverse covariance weighting improves optimality significantly. We describe optimal modal estimation in general terms applicable to polyspectra of any order in a formalism that can be efficiently extended beyond the isotropic case. We also discuss further improvements by characterising and preconditioning the covariance matrix using a single bispectrum variance shape induced by the noise and mask. We use the example of anisotropic noise in bispectrum estimation to illustrate these methods and demonstrate the necessity of subtracting a linear estimator term for optimality, removing cross-terms in the variance. We briefly discuss applications to the CMB power spectrum and trispectrum, as well as large-scale structure.

I. INTRODUCTION

A separable eigenmode approach to CMB bispectrum and trispectrum estimation has been developed and implemented [1, 2], providing a reconstruction of the WMAP CMB bispectrum and a wide variety of constraints on theoretical bispectra [3] and trispectra [4]. In this paper, we demonstrate how this modal methodology can be made fully optimal by incorporating inverse covariance weighting in the presence of anisotropic noise and other systematic effects. It can be efficiently implemented provided the eigenmode expansion utilised has sufficient resolution to describe, first, the theoretical polyspectrum shape being investigated and, secondly, the noise, mask and other systematic contributions that are correlated with it. The generality and simplicity of this modal covariance inversion, combined with the efficiency of the separable modal estimator, opens up the possibility of fully optimal analysis of high resolution CMB experiments such as the Planck satellite, as well as higher order correlators in the three-dimensional distribution of large-scale structure.

Full inverse covariance weighting for CMB power spectrum analysis has a long history going back to the study of the COBE data using direct Cholesky decomposition with $\mathcal{O}(l_{\max}^6)$ operations (see, for example, [5, 6]). This was not feasible for higher resolution experiments such as WMAP, but could be improved to $\mathcal{O}(l_{\max}^4)$ with conjugate gradient methods and iterative approximations [7]. The importance of inverse covariance weighting for the CMB bispectrum has also been discussed for some time (see, for example, [8]) but was only tackled at WMAP resolution more recently using multigrid preconditioning and conjugate gradient methods [9, 10]. Achieving full inversion at Planck resolution remains formidable goal for a realistic noise model with non-trivial correlations (see, for example, ref. [11]). The significant advance of the present work is that we show that it is not necessary to invert the full covariance matrix to obtain an optimal variance. Instead, we invert a much smaller modal covariance matrix from a subspace with sufficient degrees of freedom to describe the theoretical model, as well as the masking and noise modes that actually contaminate it. For example, we will investigate an isotropic bispectrum $b_{l_1 l_2 l_3}$ affected by anisotropic instrument noise or a cut-sky mask, discussing the projection of the pixel-variance noise map, variance and bispectrum which can affect the estimation. We have previously demonstrated, in the isotropic case, that a highly efficient and general modal bispectrum estimator can be implemented projecting the CMB data into a much smaller subspace spanned by the relevant bispectrum modes [1, 2]. What is remarkable is the small number of modes required to characterise almost all popular theoretical bispectrum models $n_{\max} = \mathcal{O}(10 - 100)$ [12]. For this isotropic subspace, we then noted that the required mode number is not

noticeably increased when characterising the relevant noise and mask effects [2]. Our expectation, therefore, when seeking an isotropic theoretical polyspectra estimator on the full anisotropic space is that we will require few additional modes to incorporate relevant noise and masking effects (this will be the subject of a future publication). Returning to the full isotropic implementation in a WMAP-realistic context, we are able to significantly improve the optimality of the variance using a much simpler inverse modal covariance weighting, getting about 30% closer to that predicted by a Fisher matrix analysis than previous isotropic bispectrum estimators.

Finally we comment on other modal estimators such as wavelets [13], bins [14] and needlets [15] which differ in their choice of initial basis but are otherwise identical. Important distinctions remain, including the ability to analytically calculate modes, the number of modes required for convergence and the ability to project between primordial and CMB(or LSS)-spectra. Early versions of the wavelet method [13], claimed optimality while neglecting to incorporate the linear term in the estimator which is required to minimise the variance [16], though this has now been incorporated [17]. We demonstrate the crucial role of this linear term for optimal estimation in the realistic situation where rotational invariance is broken.

The plan of the paper is as follows. We will first outline a general polyspectrum estimation methodology describing projections from the full polyspectrum space into the modal subspace of interest. In this context we show how the full covariance matrix is projected down into a much smaller modal covariance matrix which can be easily inverted. Next we discuss the modal bispectrum estimator in detail, deriving the relationship between the full and modal covariance matrices. WMAP results are presented comparing analytic optimal Fisher matrix forecasts with the variance obtained from Gaussian simulations for the general isotropic modal estimator. We calculate the bispectrum variance from anisotropic noise and we note that identifying the primary bispectrum noise shape contributing to this variance can be used to efficiently precondition the covariance matrix for inversion. Finally, we give the modal estimation components required for optimal CMB power spectrum and trispectrum analysis and briefly discuss large-scale structure correlators.

II. MODAL CMB POLYSPECTRA ESTIMATION

The CMB polyspectrum $\langle \mathbf{a}_\varphi \rangle$ of degree p for a given non-Gaussian model is the expectation value of the product of p temperature multipoles a_{lm} 's obtained from an ensemble of universes, that is,

$$\langle \mathbf{a}_\varphi \rangle \equiv \langle a_{l_1 m_1} a_{l_2 m_2} \dots a_{l_p m_p} \rangle \quad (1)$$

where the index φ represents the $2p$ degrees of freedom $\varphi = \{l_1, m_1, l_2, m_2, \dots, l_p, m_p\}$ which cover the full domain \mathcal{V} of possible polyspectra (up to a given resolution $l \leq l_{\max}$). Here we assume $\langle \mathbf{a}_\varphi \rangle$ is the connected part of the correlator we seek. Most primordial theories predict an isotropic polyspectrum a_ℓ which resides on a smaller d -dimensional subspace \mathcal{V}_1 , after integrating out (or summing over) the irrelevant anisotropic modes. The isotropic a_ℓ has an equivalent counterpart in the full space \mathcal{V} found by multiplying by $\mathcal{G}_{\ell\ell^\perp}$

$$\langle \mathbf{a}_\varphi \rangle = \mathcal{G}_{\ell\ell^\perp} \langle a_\ell \rangle, \quad (2)$$

where $\varphi = \{\ell, \ell^\perp\}$ with the ℓ index covering the d isotropic degrees of freedom and ℓ^\perp the remaining $2p-d$ anisotropic degrees. For concreteness, we note that the power spectrum C_l has $d = 1$ with $\mathcal{G}_{\ell\ell^\perp} = \int d\Omega_x Y_{l_1 m_1}(\hat{\mathbf{x}}) Y_{l_2 m_2}^*(\hat{\mathbf{x}}) = \delta_{l_1 l_2} \delta_{m_1 m_2}$, as we work with $a_{lm} a_{lm}^*$, while for the bispectrum we have $d = 3$ with the Gaunt integral $\mathcal{G}_{\ell\ell^\perp} = \mathcal{G}_{m_1 m_2 m_3}^{l_1 l_2 l_3} \equiv \int d\Omega_x Y_{l_1 m_1}(\hat{\mathbf{x}}) Y_{l_2 m_2}(\hat{\mathbf{x}}) Y_{l_3 m_3}(\hat{\mathbf{x}})$, $\ell \rightarrow \{l_1, l_2, l_3\}$, $\ell^\perp \rightarrow \{m_1, m_2, m_3\}$ (and similar for higher ‘nondiagonal’ polyspectra).

When investigating CMB polyspectra we are typically seeking to match a statistically isotropic prediction \mathbf{a}_{φ_i} to the CMB data, with an appropriate weighting to reflect our knowledge of the signal-to-noise for the specific experiment. The estimator $\bar{\mathcal{E}}$ determines the goodness of fit between these through the inner product

$$\bar{\mathcal{E}} \equiv \frac{\sum_{\varphi\varphi'} \langle \mathbf{a}_\varphi \rangle \mathfrak{C}_{\varphi\varphi'}^{-1} (\mathbf{a}_\varphi - \mathbf{a}_\varphi^{lin})}{\sum_{\varphi\varphi'} \langle \mathbf{a}_\varphi \rangle \mathfrak{C}_{\varphi\varphi'}^{-1} \langle \mathbf{a}_\varphi \rangle}, \quad (3)$$

where the $\mathfrak{C}_{\varphi\varphi'}^{-1} \equiv C_{l_1 m_1, l'_1 m'_1}^{-1} \dots C_{l_p m_p, l'_p m'_p}^{-1}$ represents the inverse covariance across all p multipoles. Here, $\mathfrak{a}_{\varphi}^{lin}$ represents the ‘linear term’ used to subtract out known systematic effects; for the power spectrum, $\mathfrak{a}_{\varphi}^{lin} = C_{l_1 m_1, l_2 m_2}$ and for the bispectrum $\mathfrak{a}_{\varphi}^{lin} = 3C_{l_1 m_1, l_2 m_2} a_{l_3 m_3}$. In practice, the linear term is calculated from Monte Carlo simulations of Gaussian maps with the correct experimental effects (mask, beam, noise) included, see [2]. We also note that $\langle \mathfrak{a}_{\varphi} \rangle$ is understood to be the theoretical prediction modified to include the relevant experimental effects.

If we define the weighted vectors \mathcal{A} , \mathcal{B} and the matrix \mathcal{C}

$$\mathcal{A}_{\varphi} = \frac{\langle \mathfrak{a}_{\varphi} \rangle}{\sqrt{C_{l_1} C_{l_2} \dots C_{l_p}}}, \quad \mathcal{B}_{\varphi} = \frac{\mathfrak{a}_{\varphi} - \mathfrak{a}_{\varphi}^{lin}}{\sqrt{C_{l_1} C_{l_2} \dots C_{l_p}}}, \quad \mathcal{C}_{\varphi\varphi'} = \frac{\mathfrak{C}_{\varphi\varphi'}}{\sqrt{C_{l_1} C_{l'_1} \dots C_{l_p} C_{l'_p}}}, \quad (4)$$

then the estimator can be written in matrix notation as

$$\bar{\mathcal{E}} = \frac{\mathcal{A}^T \mathcal{C}^{-1} \mathcal{B}}{\mathcal{A}^T \mathcal{C}^{-1} \mathcal{A}}. \quad (5)$$

Modal decompositions have proven to be very useful for estimating general CMB bispectra and trispectra. Suppose we have a basis $\mathcal{R}_{n\mathfrak{p}}$ where n is the mode number. We will take it to be orthonormal so

$$\sum_{\mathfrak{p}} \mathcal{R}_{n\mathfrak{p}} \mathcal{R}_{n'\mathfrak{p}} = \delta_{nn'}, \quad (6)$$

or in matrix notation $\mathcal{R}\mathcal{R}^T = I$. We can then represent our theoretical polyspectra in this basis using mode coefficients α defined by

$$\mathcal{A}_{\mathfrak{p}} = \sum_n \alpha_n \mathcal{R}_{n\mathfrak{p}} \quad (\mathcal{A} = \mathcal{R}^T \alpha). \quad (7)$$

The α can then be calculated easily via $\alpha = \mathcal{R}\mathcal{A}$.

Most theoretical cosmological models have rotational invariance at their foundation, so the polyspectra \mathcal{A}^{th} they predict are generically isotropic. For this reason, our starting point must be to include sufficient isotropic mode functions in our basis to accurately characterise them. Most theoretical bispectra \mathcal{A}^{th} have smooth or simple forms and so for a suitable choice of \mathcal{R} they can in fact be represented by surprisingly few of these isotropic modes (for example, most scale-invariant bispectrum models require only 15 modes at WMAP resolution [12]). This motivates us to work with a truncated basis containing n_{max} modes which span the space of interest which is determined by the class of theoretical models under consideration modified by experimental effects, i.e. masking and noise. For an isotropic bispectrum estimator, we have shown that the experimental effects of the mask and noise can be incorporated without increasing n_{max} substantially [2]. For anisotropic contributions from experimental effects which yield a partially off-diagonal \mathcal{A} , we need to supplement our basis functions with additional anisotropic modes. In line with the remarkable data compression possible in the isotropic case, we anticipate only increasing n_{max} by multiplying by a factor of a few provided the extra anisotropic \mathcal{R} are carefully chosen.

Given that n_{max} is so small relative to the vast number $2p$ of multipole modes, the \mathcal{R} become highly rectangular $n_{max} \times 2p$ matrices. We can now define a projection operator from the full space \mathcal{V} to the subspace $\mathcal{V}_{\mathcal{P}}$

$$\mathcal{P} \equiv \mathcal{R}^T \mathcal{R} \quad (8)$$

Here, we assume we have truncated our basis such that our theory and distorting experimental effects remain accurately described, ie. $\mathcal{P}\mathcal{A} = \mathcal{A}$. Using this projection we define the modal counterparts of the data on the same subspace $\mathcal{V}_{\mathcal{P}}$,

$$\beta = \mathcal{R}\mathcal{B} \longrightarrow \mathcal{P}\mathcal{B} = \mathcal{R}^T \beta, \quad (9)$$

as well as the covariance

$$\zeta = \mathcal{R}\mathcal{C}\mathcal{R}^T. \quad (10)$$

The projected covariance is related to the covariance of β by $p!$ as we now show.

$$\langle \beta \beta^T \rangle = \mathcal{R} \langle \mathcal{B} \mathcal{B}^T \rangle \mathcal{R}^T \quad (11)$$

$$= \mathcal{R} \frac{\langle \mathbf{a}_\varphi \mathbf{a}_{\varphi'} \rangle - \langle \mathbf{a}_\varphi \mathbf{a}_{\varphi'}^{lin} \rangle - \langle \mathbf{a}_\varphi^{lin} \mathbf{a}_{\varphi'} \rangle + \langle \mathbf{a}_\varphi^{lin} \mathbf{a}_{\varphi'}^{lin} \rangle}{\sqrt{C_{l_1} \dots C_{l_p} C_{l'_1} \dots C_{l'_p}}} \mathcal{R}^T. \quad (12)$$

As we know the a_{lm} are very close to Gaussian we can expand $\langle \mathbf{a}_\varphi \mathbf{a}_{\varphi'} \rangle$ into the product $\langle a_{lm} a_{l'm'} \rangle, \langle a_{lm} a_{lm} \rangle$ and $\langle a_{l'm'} a_{l'm'} \rangle$. In general there will be $p!$ ways to have all pairs consist of one primed and one unprimed a_{lm} plus $(2p)!/2p! - p!$ combinations with at least one pairing between two unprimed and two primed a_{lm} . The linear term is always chosen such that it removes all contributions of the second type. As a result we have $\langle \mathcal{B} \mathcal{B}^T \rangle = p! \mathcal{C}$ and so

$$\zeta = \mathcal{R}\mathcal{C}\mathcal{R}^T = \frac{1}{p!} \langle \beta \beta^T \rangle \quad (13)$$

Using these quantities we can define a new estimator

$$\mathcal{E} \equiv \frac{\alpha^T \zeta^{-1} \beta}{\alpha^T \zeta^{-1} \alpha} \quad (14)$$

$$\begin{aligned} &= \frac{(\mathcal{R}\mathcal{A})^T \mathcal{R} \mathcal{C}^{-1} \mathcal{R}^T \mathcal{R} \mathcal{B}}{\mathcal{R} \mathcal{A}^T \mathcal{R} \mathcal{C}^{-1} \mathcal{R}^T \mathcal{R} \mathcal{A}} \\ &= \frac{\mathcal{A} \mathcal{P} \mathcal{C}^{-1} \mathcal{P} \mathcal{B}}{\mathcal{A}^T \mathcal{P} \mathcal{C}^{-1} \mathcal{P} \mathcal{A}} \end{aligned} \quad (15)$$

where we have used $(\mathcal{R}\mathcal{C}\mathcal{R}^T)^{-1} = \mathcal{R}\mathcal{C}^{-1}\mathcal{R}^T + Z_\perp$ where Z_\perp is an arbitrary term orthogonal to the projection which we can ignore. Strictly \mathcal{C} is non-invertible, formally because it contains duplicate data due to permutations, but also because of loss of information from a cut-sky mask. However, this problem can be circumvented. For example, we can restrict φ to count only combinations where $l_1 \leq l_2 \leq l_3$, with a suitable restriction on m_i when the l_i are equal, and weight them appropriately to remove this complication. Once it is inverted we can then reinsert duplicates to return to the previous ordering for φ . Given the estimator expression (14), it can be easily seen that its variance is given by $\sigma^2 \propto (\alpha^T \zeta^{-1} \alpha)^{-1}$ since $\zeta \propto \langle \beta \beta^T \rangle$.

We now wish to determine how closely the standard estimator (5) and the modal estimator (14) coincide. In the idealised case of homogeneous noise and no mask the covariance matrix is diagonal and so $\mathcal{C} = I$ and they are clearly identical. Also in the limit of completeness $n_{max} \rightarrow 2p$, we have $\mathcal{P} \rightarrow I$ and again the estimators are equivalent. In a realistic context with inhomogeneous noise, a mask and a truncated set of modes the difference can be quantified by considering

$$\mathcal{A} = \begin{bmatrix} \mathcal{A}_\parallel \\ 0 \end{bmatrix} \quad \mathcal{B} = \begin{bmatrix} \mathcal{B}_\parallel \\ \mathcal{B}_\perp \end{bmatrix} \quad \mathcal{C}^{-1} = \begin{bmatrix} \mathcal{C}_\parallel^{-1} & \mathcal{C}_\times^{-1} \\ \mathcal{C}_{\times}^{-1T} & \mathcal{C}_\perp^{-1} \end{bmatrix} \quad (16)$$

where we define the vector $\mathcal{X}_\parallel \equiv \mathcal{P}\mathcal{X}$ to be the component of \mathcal{X} in the subspace $\mathcal{V}_\mathcal{P}$ and the vector $\mathcal{X}_\perp \equiv (I - \mathcal{P})\mathcal{X}$ to be orthogonal to it and, similarly, for the matrix we can decompose it into $\mathcal{M}_\parallel \equiv \mathcal{P}\mathcal{M}\mathcal{P}$, $\mathcal{M}_\perp \equiv (I - \mathcal{P})\mathcal{M}(I - \mathcal{P})$ and $\mathcal{M}_\times \equiv \mathcal{P}\mathcal{M}(I - \mathcal{P})$. In this case, we can represent the standard estimator (5) in the form,

$$\bar{\mathcal{E}} = \frac{\mathcal{A}_\parallel (\mathcal{C}_\parallel^{-1} \mathcal{B}_\parallel + \mathcal{C}_\times^{-1} \mathcal{B}_\perp)}{\mathcal{A}_\parallel^T \mathcal{C}_\parallel^{-1} \mathcal{A}_\parallel}, \quad (17)$$

whereas our new modal estimator (14) is

$$\mathcal{E} = \frac{\mathcal{A}_{\parallel} \mathcal{C}_{\parallel}^{-1} \mathcal{B}_{\parallel}}{\mathcal{A}_{\parallel}^T \mathcal{C}_{\parallel}^{-1} \mathcal{A}_{\parallel}}. \quad (18)$$

The difference between the estimators is the inverse covariance projection of the perpendicular part of the data onto the parallel space. We should however pause to note that the normalisation in both cases is identical. As the variance we expect for the estimator is the reciprocal of the normalisation (with a factor for permutations), we can see the effect of the cross-projection by comparing the variance we obtain from simulations with this prediction. If they differ we must deduce that our covariance weighting is sub-optimal and the cross-term is important in which case we must increase n_{max} and expand our basis to include the relevant parts.

It should be noted that as the covariance matrix ζ is symmetric positive definite it is possible to find its Cholesky decomposition,

$$\zeta = \tilde{\lambda} \tilde{\lambda}^T. \quad (19)$$

where $\tilde{\lambda}$ is lower triangular. We can then absorb these factors into our α and β ,

$$\alpha' = \tilde{\lambda}^{-1} \alpha \quad \beta' = \tilde{\lambda}^{-1} \beta. \quad (20)$$

This is equivalent to re-orthonormalising our basis with respect to the full covariance matrix, and hence the β are now uncorrelated. This is obviously a desirable property and has several applications which we explore elsewhere in future publications. The most obvious is that the estimator takes a very simple form

$$\mathcal{E} = \frac{\alpha'^T \beta'}{\alpha'^T \alpha'} \quad (21)$$

as by construction $\zeta' = I$.

The computational advantages of the modal estimator (14) are considerable, especially when we note that direct evaluation of the full optimal estimator (5) is extremely challenging at high resolution. This is because, first, it represents a huge sum over l_{max}^{2p} quantities and, secondly, the inverse covariance weighting requires a huge $l_{max}^2 \times l_{max}^2$ matrix to be calculated and then inverted. As we noted in the introduction, there do exist methods which make this tractable in specific cases, however for large l_{max} and general noise models at Planck resolution this has not yet been achieved. Calculation of the mode coefficients α, β may appear to be troublesome, but it is made highly efficient using a separable form which has been discussed at length previously [2]. Instead, here, we focus on the improvements offered by the modal approach for inverse covariance weighting, reducing this to the inversion of a small $n_{max} \times n_{max}$ matrix.

III. CMB BISPECTRUM ESTIMATION

A. Inverse covariance weighting with the modal bispectrum estimator

Using the CMB bispectrum, we will now demonstrate a concrete example and implementation of modal polyspectra estimation with inverse covariance weighting. The estimator for the bispectrum takes the general form

$$\mathcal{E} = \frac{1}{N} \sum_{l_i m_i} B_{m_1 m_2 m_3}^{l_1 l_2 l_3} C_{l_1 l'_1 m_1 m'_1}^{-1} C_{l_2 l'_2 m_2 m'_2}^{-1} C_{l_3 l'_3 m_3 m'_3}^{-1} (a_{l'_1 m'_1} a_{l'_2 m'_2} a_{l'_3 m'_3} - 3 C_{l'_1 l'_2 m'_1 m'_2} a_{l'_3 m'_3}) , \quad (22)$$

where $B_{m_1 m_2 m_3}^{l_1 l_2 l_3} = \langle a_{l_1 m_1} a_{l_2 m_2} a_{l_3 m_3} \rangle$ is the full bispectrum, modified from being purely isotropic by experimental effects, and N is the appropriate normalisation. This estimator has been shown to be optimal in the ideal case without mask or noise in the absence of a linear term [18]. The linear term was proposed in

[16] to minimise the variance in a realistic setting when rotational invariance is broken. To construct the modal bispectrum estimator, the quantities $\mathcal{A}, \mathcal{B}, \mathcal{C}$ in (4) take the form

$$\mathcal{A} = \frac{B_{m_1 m_2 m_3}^{l_1 l_2 l_3}}{\sqrt{C_{l_1} C_{l_2} C_{l_3}}}, \quad (23)$$

$$\mathcal{B} = \frac{a_{l_1 m_1} a_{l_2 m_2} a_{l_3 m_3} - 3 C_{l_1 m_1, l_2 m_2} a_{l_3 m_3}}{\sqrt{C_{l_1} C_{l_2} C_{l_3}}}, \quad (24)$$

$$\mathcal{C} = \frac{C_{l_1 m_1, l'_1 m'_1} C_{l_2 m_2, l'_2 m'_2} C_{l_3 m_3, l'_3 m'_3}}{\sqrt{C_{l_1} C_{l_2} C_{l_3} C_{l'_1} C_{l'_2} C_{l'_3}}}. \quad (25)$$

For simplicity and to relate the discussion more easily to previous work, let us restrict attention to an isotropic subspace, though the expressions derived below can be converted for the general anisotropic case. We shall take the usual approximation that the primary contribution to the isotropic bispectrum is the predicted theoretical model itself, that is,

$$\mathcal{A} = \mathcal{G}_{m_1 m_2 m_3}^{l_1 l_2 l_3} \frac{b_{l_1 l_2 l_3}}{\sqrt{C_{l_1} C_{l_2} C_{l_3}}} \quad (26)$$

where $b_{l_1 l_2 l_3}$ is the predicted reduced bispectrum. Here we note that we are always working with quantities that include the relevant experimental effects. For example, as we are only considering the isotropic case we can approximate $b_{l_1 l_2 l_3} \approx f_{sky} b_{l_1} b_{l_2} b_{l_3} b_{l_1 l_2 l_3}^{clean}$ and $C_l \approx f_{sky} (b_l^2 C_l^{clean} + N_l)$ where b_l is the beam window function, N_l is the noise power spectrum and f_{sky} is the fraction of the sky remaining after masking.

An isotropic set of basis functions can be represented in the full anisotropic bispectrum space by

$$\mathcal{R}_{m_1 m_2 m_3}^{l_1 l_2 l_3} = \frac{\mathcal{G}_{m_1 m_2 m_3}^{l_1 l_2 l_3}}{v_{l_1} v_{l_2} v_{l_3}} R_{nl_1 l_2 l_3}, \quad (27)$$

where v_l is a weight function chosen to mimic the scaling of \mathcal{G} , the Gaunt integral. The simple orthogonality condition $\mathcal{R} \mathcal{R}^T = I$ in the full space \mathcal{V} , yields a weighted orthogonality condition on R in the projected isotropic subspace \mathcal{V}_P ,

$$\delta_{nn'} = \sum_{l_i m_i} \frac{(\mathcal{G}_{m_1 m_2 m_3}^{l_1 l_2 l_3})^2}{v_{l_1}^2 v_{l_2}^2 v_{l_3}^2} R_{nl_1 l_2 l_3} R_{n' l_1 l_2 l_3} = \sum_{l_i} \frac{h_{l_1 l_2 l_3}^2}{v_{l_1}^2 v_{l_2}^2 v_{l_3}^2} R_{nl_1 l_2 l_3} R_{n' l_1 l_2 l_3} \quad (28)$$

where $h_{l_1 l_2 l_3}$ is a geometric factor defined by

$$h_{l_1 l_2 l_3} = \sqrt{\frac{(2l_1 + 1)(2l_2 + 1)(2l_3 + 1)}{4\pi}} \begin{pmatrix} l_1 & l_2 & l_3 \\ 0 & 0 & 0 \end{pmatrix} \quad (29)$$

and decomposing the $\mathcal{A}, \mathcal{B}, \mathcal{C}$ into mode space yields the α_n, β_n coefficients and the covariance matrix $\zeta_{nn'}$ respectively, that is,

$$\alpha_n = \sum_{l_i} \frac{(2l_1 + 1)(2l_2 + 1)(2l_3 + 1)}{4\pi v_{l_1}^2 v_{l_2}^2 v_{l_3}^2} \begin{pmatrix} l_1 & l_2 & l_3 \\ 0 & 0 & 0 \end{pmatrix}^2 \frac{v_{l_1} v_{l_2} v_{l_3} b_{l_1 l_2 l_3}}{\sqrt{C_{l_1} C_{l_2} C_{l_3}}} R_{nl_1 l_2 l_3} \quad (30)$$

$$\beta_n = \sum_{l_i m_i} \frac{\mathcal{G}_{m_1 m_2 m_3}^{l_1 l_2 l_3}}{v_{l_1} v_{l_2} v_{l_3}} \frac{a_{l_1 m_1} a_{l_2 m_2} a_{l_3 m_3} - 3 C_{l_1 m_1, l_2 m_2} a_{l_3 m_3}}{\sqrt{C_{l_1} C_{l_2} C_{l_3}}} R_{nl_1 l_2 l_3} \quad (31)$$

$$\zeta_{nn'} = \sum_{l_i m_i l'_i m'_i} \frac{\mathcal{G}_{m_1 m_2 m_3}^{l_1 l_2 l_3} \mathcal{G}_{m'_1 m'_2 m'_3}^{l'_1 l'_2 l'_3}}{v_{l_1} v_{l_2} v_{l_3} v_{l'_1} v_{l'_2} v_{l'_3}} R_{nl_1 l_2 l_3} \frac{C_{l_1 m_1, l'_1 m'_1} C_{l_2 m_2, l'_2 m'_2} C_{l_3 m_3, l'_3 m'_3}}{\sqrt{C_{l_1} C_{l_2} C_{l_3} C_{l'_1} C_{l'_2} C_{l'_3}}} R_{n' l'_1 l'_2 l'_3}. \quad (32)$$

Calculation of α_n , β_n is straightforward using methods described in detail in ref. [2] and, as we have chosen the linear term correctly to remove cross terms in the variance, $\zeta = 1/6 \langle \beta \beta^R \rangle$ as we demonstrate below

$$\langle \beta_n \beta_{n'} \rangle = \sum_{l_i m_i l'_i m'_i} \left\langle \left(\frac{\mathcal{G}_{m_1 m_2 m_3}^{l_1 l_2 l_3} a_{l_1 m_1} a_{l_2 m_2} a_{l_3 m_3} - 3 C_{l_1 m_1, l_2 m_2} a_{l_3 m_3} R_{nl_1 l_2 l_3}}{v_{l_1} v_{l_2} v_{l_3} \sqrt{C_{l_1} C_{l_2} C_{l_3}}} \right) \right. \\ \left. \times \left(\frac{\mathcal{G}_{m'_1 m'_2 m'_3}^{l'_1 l'_2 l'_3} a_{l'_1 m'_1} a_{l'_2 m'_2} a_{l'_3 m'_3} - 3 C_{l'_1 m'_1, l'_2 m'_2} a_{l'_3 m'_3} R_{nl'_1 l'_2 l'_3}}{v_{l'_1} v_{l'_2} v_{l'_3} \sqrt{C_{l'_1} C_{l'_2} C_{l'_3}}} \right) \right\rangle \quad (33)$$

$$= \sum_{l_i m_i l'_i m'_i} \frac{\mathcal{G}_{m_1 m_2 m_3}^{l_1 l_2 l_3} \mathcal{G}_{m'_1 m'_2 m'_3}^{l'_1 l'_2 l'_3}}{v_{l_1} v_{l_2} v_{l_3} v_{l'_1} v_{l'_2} v_{l'_3}} R_{nl_1 l_2 l_3} [6 \langle a_{l_1 m_1} a_{l'_1 m'_1} \rangle \langle a_{l_2 m_2} a_{l'_2 m'_2} \rangle \langle a_{l_3 m_3} a_{l'_3 m'_3} \rangle \\ + 9 \langle a_{l_1 m_1} a_{l_2 m_2} \rangle \langle a_{l'_1 m'_1} a_{l'_2 m'_2} \rangle \langle a_{l_3 m_3} a_{l'_3 m'_3} \rangle - 9 C_{l_1 m_1, l_2 m_2} \langle a_{l'_1 m'_1} a_{l'_2 m'_2} \rangle \langle a_{l_3 m_3} a_{l'_3 m'_3} \rangle \\ - 9 \langle a_{l_1 m_1} a_{l_2 m_2} \rangle C_{l'_1 m'_1, l'_2 m'_2} \langle a_{l_3 m_3} a_{l'_3 m'_3} \rangle + 9 C_{l_1 m_1, l_2 m_2} C_{l'_1 m'_1, l'_2 m'_2} \langle a_{l_3 m_3} a_{l'_3 m'_3} \rangle + \dots] R_{nl'_1 l'_2 l'_3} \\ = 6 \sum_{l_i m_i l'_i m'_i} \frac{\mathcal{G}_{m_1 m_2 m_3}^{l_1 l_2 l_3} \mathcal{G}_{m'_1 m'_2 m'_3}^{l'_1 l'_2 l'_3}}{v_{l_1} v_{l_2} v_{l_3} v_{l'_1} v_{l'_2} v_{l'_3}} R_{nl_1 l_2 l_3} \frac{C_{l_1 m_1, l'_1 m'_1} C_{l_2 m_2, l'_2 m'_2} C_{l_3 m_3, l'_3 m'_3}}{\sqrt{C_{l_1} C_{l_2} C_{l_3} C_{l'_1} C_{l'_2} C_{l'_3}}} R_{nl'_1 l'_2 l'_3} + \text{higher order} \\ = 6 \mathcal{R} \mathcal{C} \mathcal{R}^T = 6 \zeta_{nn'} . \quad (34) \quad (35)$$

Our inverse covariance weighted estimator (18) becomes

$$\mathcal{E} = \frac{\alpha_n \zeta_{nn'}^{-1} \beta_{n'}}{\alpha_n \zeta_{nn'}^{-1} \alpha_{n'}} . \quad (36)$$

This result applies equally to the general anisotropic case where we have supplemented with basis functions $\mathcal{R}_{m_1 m_2 m_3}^{l_1 l_2 l_3}$ which cannot be simply represented in terms of the Gaunt integral; the derivation above is unaffected. Using (36), we note that the variance σ^2 of the optimal modal estimator (36) becomes simply

$$\sigma^2 = \frac{\alpha_n \zeta_{nn'}^{-1} \langle \beta_{n'} \beta_{p'} \rangle \zeta_{pp'}^{-1} \alpha_p}{(\alpha_n \zeta_{nn'}^{-1} \alpha_{n'})^2} = \frac{6}{\alpha_n \zeta_{nn'}^{-1} \alpha_{n'}} . \quad (37)$$

B. Bivector representation of covariance matrix

Further efficiencies are made possible by identifying the underlying bispectrum ‘noise’ and ‘mask’ shapes that contribute to the covariance matrix. We can identify such a shape because the noise is correlated in harmonic space. Empirically in our previous isotropic WMAP analysis we have found that the noise and mask contribute in tandem to produce a single shape vector $\beta = \{\tilde{\beta}_n\}$ which we could identify [3]. This noise/mask vector is significantly correlated with local non-Gaussianity which explains the difficulty in obtaining optimal constraints on this model. The dominant effect of β leads to the remarkable result that we can accurately represent the isotropic covariance matrix simply as the identity matrix perturbed by this bivector, that is,

$$\zeta_{nn'} \approx \delta_{nn'} + \tilde{\beta}_n \tilde{\beta}_{n'} . \quad (38)$$

For this, there is a simple inverse covariance

$$\zeta_{nn'}^{-1} \approx \delta_{nn'} - \frac{\tilde{\beta}_n \tilde{\beta}_{n'}}{1 + |\tilde{\beta}|^2}, \quad \text{where } |\tilde{\beta}|^2 \equiv \sum_{i=1}^{n_{\max}} \tilde{\beta}_i^2 . \quad (39)$$

For the isotropic estimator (36) in the presence of this noise/mask bivector the variance from (37) becomes

$$\begin{aligned}\sigma^2 &= \frac{1}{|\alpha|^2} \left(1 + \frac{(\alpha \cdot \tilde{\beta})^2}{|\alpha|^2(1 + |\tilde{\beta}|^2) - (\alpha \cdot \tilde{\beta})^2} \right) \\ &= \frac{1}{|\alpha|^2} \left(1 + \frac{|\tilde{\beta}|^2 \cos^2 \theta}{1 + |\tilde{\beta}|^2 (1 - \cos^2 \theta)} \right),\end{aligned}\quad (40)$$

where we have described the degree of correlation between the two vector directions through the cosine

$$\alpha \cdot \tilde{\beta} = \sum_{i=1}^{n_{\max}} \alpha_n \tilde{\beta}_n \equiv |\alpha| |\tilde{\beta}| \cos \theta. \quad (41)$$

From (40), we see that the variance is made up of two components. First, there is the ideal or ‘optimal’ variance $1/|\alpha|^2$ that can be achieved in the presence of the underlying Gaussian CMB signal and homogenous instrument noise, given the degree of sky coverage and beam resolution (incorporated into the signal-to-noise definition of α). Secondly, there is the contribution from the inhomogeneous noise and the anisotropic mask which arises if the noise/mask vector $\tilde{\beta}$ is closely aligned with the model α under investigation. Significantly, this second contribution vanishes if the model being studied has little correlation with $\tilde{\beta}$, in which case near-optimal variance can be approached even in this isotropic case. In principle, in the full anisotropic case, the general inverse covariance weighting will approach optimality given sufficient modes for any model.

Another key point in a practical implementation is that the single noise/mask vector $\tilde{\beta}$ can be estimated accurately from relatively few simulations. Using $\tilde{\beta}$ with the covariance matrix in the form (38) allows accurate and efficient inversion because we only need to determine $\tilde{\beta}$. This can be achieved with $n_{\text{sim}} \lesssim 1000$ simulations, while direct averaging of the covariance matrix takes $n_{\text{sim}} \approx 10000$ to converge to percent levels (making it reliably invertible). We remark that early studies of these contributions in a WMAP-realistic context showed that the local shape is strongly correlated with the spurious contributions from the mask and inhomogenous noise, with the noise/mask vector $\tilde{\beta}$ showing an 70% correlation with the local α^{loc} direction. Separately, there was about an 70% local-mask correlation and a 47% local-noise correlation. In a forthcoming publication, we shall discuss the more general case where several ‘noise’ vectors can be obtained using a principal component analysis (PCA), separating the noise, mask and other contaminant vectors from model directions of theoretical interest.

Finally, we note that for the simple bivector form of the inverse covariance matrix (39), the Cholesky decomposition of ζ can also be expressed analytically. Here, we have $\zeta = \tilde{\lambda} \tilde{\lambda}^T$ with $\tilde{\lambda}$ the lower triangular matrix:

$$\tilde{\lambda} = \begin{pmatrix} \sqrt{1 + \tilde{\beta}_1^2} & 0 & 0 & 0 \\ \frac{\tilde{\beta}_2 \tilde{\beta}_1}{\sqrt{1 + \tilde{\beta}_1^2}} & \frac{\sqrt{1 + \tilde{\beta}_1^2 + \tilde{\beta}_2^2}}{\sqrt{1 + \tilde{\beta}_1^2}} & 0 & 0 \\ \dots & \dots & \dots & \dots \\ \frac{\tilde{\beta}_n \tilde{\beta}_1}{\sqrt{1 + \tilde{\beta}_1^2}} & \dots & \frac{\tilde{\beta}_n \tilde{\beta}_{n'}}{\sqrt{1 + \sum_1^{n'-1} \tilde{\beta}_n^2} \sqrt{1 + \sum_1^{n'} \tilde{\beta}_i^2}} & \frac{\sqrt{1 + \sum_1^n \tilde{\beta}_i^2}}{\sqrt{1 + \sum_1^{n-1} \tilde{\beta}_i^2}} \end{pmatrix}. \quad (42)$$

Rotation with this matrix $\tilde{\lambda}$ using (20), offers an orthonormal frame in which the bispectrum estimator (36) becomes simply $\mathcal{E} = \alpha' \cdot \beta' / |\alpha'|^2$, that is, the effects of the noise and mask are incorporated into the definitions of α' and β' from the outset.

C. Isotropic inverse covariance implementation

The practical implementation of the inverse covariance weighting in (18) has been incorporated in the isotropic case for the modal bispectrum estimator. Bispectrum constraints for a wide variety of models using WMAP data have been presented elsewhere [3]. Here, our purpose is to focus on the improvement

in optimality for the variance Δf_{NL} and we only present results for local, equilateral and orthogonal non-Gaussian models. We compare the variance achieved for the modal isotropic estimator (which agrees well with previous isotropic KSW estimators) to show that it can be improved with modal inverse covariance weighting in this isotropic case. Of course, we do not achieve full optimality, which can be seen by comparing to the Fisher matrix forecast, because relevant anisotropic effects deriving chiefly from the mask cannot be described using only an isotropic basis.

For the purposes of direct comparison, we present the variance ΔF_{NL} using the integrated bispectrum F_{NL} defined in ref. [3], that is, the total bispectrum signal-to-noise normalised relative to the local model (for which $f_{\text{NL}}^{\text{loc}} = F_{\text{NL}}$). For an arbitrary predicted theoretical bispectrum $b_{l_1 l_2 l_3}^{\text{th}}$, we use the estimator (22) to define

$$F_{\text{NL}}^{\text{th}} = \mathcal{E}, \quad \text{with} \quad N^2 \equiv N_{\text{th}} N_{\text{loc}}^{f_{\text{NL}}=1}, \quad (43)$$

where the universal normalization N has been defined using

$$N_{\text{th}}^2 \equiv \sum_{l_i} \frac{h_{l_1 l_2 l_3}^2 b_{l_1 l_2 l_3}^{\text{th}^2}}{C_{l_1} C_{l_2} C_{l_3}}. \quad (44)$$

with $N_{\text{loc}}^{f_{\text{NL}}=1}$ being exactly the same quantity except defined for the $f_{\text{NL}} = 1$ local model. The quantity F_{NL} provides a much more uniform variance between different models, though we also present results with the usual equilateral normalisation $f_{\text{NL}}^{\text{equil}}$.

Estimation type	Local ΔF_{NL}	Equilateral ΔF_{NL} ($\Delta f_{\text{NL}}^{\text{equil}}$)	Orthogonal ΔF_{NL} ($\Delta f_{\text{NL}}^{\text{ortho}}$)
Standard isotropic estimator	29.5	24.1 (129.0)	26.0 (107.9)
Weighted isotropic estimator	27.6	23.8 (127.4)	25.0 (103.8)
Fisher matrix ‘optimal’	22.9	22.9 (122.5)	22.9 (94.9)

Table I: Results for the variance from an implementation of the isotropic modal estimator in a WMAP realistic context. These results were obtained from 1000 Gaussian maps at $l_{\text{max}} = 500$ using the KQ75 mask and the WMAP inhomogeneous noise model. The mask and noise degrades the variance relative to the Fisher matrix forecast, but this is significantly improved using modal inverse covariance weighting, even for this isotropic estimator.

Results for the variance ΔF_{NL} from local and equilateral models are listed in Table I. This work conforms exactly to the WMAP-realistic methods described in ref. [3], using a KQ75 mask, WMAP beams and inhomogeneous noise model (see below) and with a resolution limit of $l_{\text{max}} = 500$. The results were obtained using 144000 Gaussian maps from a modal bispectrum pipeline previously confirmed to be unbiased and which reproduces the correct f_{NL} for simulated non-Gaussian maps [2]. Previous results for ideal full-sky Gaussian simulations (without inhomogeneous noise or mask) demonstrate that the estimator is correctly normalised, confirming that optimality variance in the absence of experimental effects. For the local model in a WMAP-realistic context, the variance from the modal estimator is suboptimal by 30% dropping from the Fisher matrix forecast $\Delta F_{\text{NL}} = 22.9$ to a realised value $\Delta F_{\text{NL}} = 29.5$. This is consistent with the value obtained from KSW estimators previously at the same $l_{\text{max}} = 500$ (e.g. $\Delta F_{\text{NL}} \approx 26$ in ref. [19]). This strong degradation for the local model occurs because the noise/mask vector (38)) correlates strongly with the local shape, increasing the variance as discussed previously (see (40)). The addition of inverse covariance weighting for the local model, improves the variance to $\Delta F_{\text{NL}} = 27.6$ which takes us 30% closer to the projected optimal variance, demonstrating the value of incorporating this weighting even in the purely isotropic case. In contrast, nearly optimal results can always be obtained for the equilateral model and the orthogonal model because these are much less correlated with the noise/mask vector $\tilde{\beta}$. For the equilateral model, in particular, an isotropic estimator will be capable of achieving near-optimal results.

IV. OTHER POLYSPECTRA AND 3D APPLICATIONS

This methodology can be extended to other correlators. We begin with the trispectrum estimator presented in [20], for which WMAP constraints were obtained for several models in [4]. We shall restrict

ourselves to the diagonal free trispectrum for simplicity, although obtaining expressions for the general case is straight forward. We build the optimal modal estimator out of the quantities \mathcal{A} , \mathcal{B} , \mathcal{C} in (4) defined by

$$\mathcal{A} = \frac{G_{m_1 m_2 m_3 m_4}^{l_1 l_2 l_3 l_4} t_{l_1 l_2 l_3 l_4}}{\sqrt{C_{l_1} C_{l_2} C_{l_3} C_{l_4}}}, \quad (45)$$

$$\mathcal{B} = \frac{a_{l_1 m_1} a_{l_2 m_2} a_{l_3 m_3} a_{l_4 m_4} - 6 C_{l_1 m_1, l_2 m_2} a_{l_3 m_3} a_{l_4 m_4} + 3 C_{l_1 m_1, l_2 m_2} C_{l_3 m_3, l_4 m_4}}{\sqrt{C_{l_1} C_{l_2} C_{l_3} C_{l_4}}}, \quad (46)$$

$$\mathcal{C} = \frac{C_{l_1 m_1, l'_1 m'_1} C_{l_2 m_2, l'_2 m'_2} C_{l_3 m_3, l'_3 m'_3} C_{l_4 m_4, l'_4 m'_4}}{\sqrt{C_{l_1} C_{l_2} C_{l_3} C_{l_4} C_{l'_1} C_{l'_2} C_{l'_3} C_{l'_4}}}, \quad (47)$$

$$\mathcal{R} = \frac{G_{m_1 m_2 m_3 m_4}^{l_1 l_2 l_3 l_4}}{v_{l_1} v_{l_2} v_{l_3} v_{l_4}} R_{nl_1 l_2 l_3 l_4}, \quad (48)$$

where v_l is a suitable weight function and the isotropising factor $G_{m_1 m_2 m_3 m_4}^{l_1 l_2 l_3 l_4}$ is

$$G_{m_1 m_2 m_3 m_4}^{l_1 l_2 l_3 l_4} = \int Y_{l_1 m_1} Y_{l_2 m_2} Y_{l_3 m_3} Y_{l_4 m_4} = \sum_{LM} (-1)^M G_{m_1 m_2 M}^{l_1 l_2 L} G_{m_3 m_4 -M}^{l_3 l_4 L}. \quad (49)$$

Using efficient modal methods calculation of the trispectrum estimator with inverse covariance weighting is no more challenging than in the bispectrum.

Inverse covariance weighting for power spectrum estimation for high resolution experiments like Planck. In this case we would be constraining small deviations from a canonical model with a given C_l . Considering an isotropic subspace, the building blocks for the model estimator are

$$\mathcal{A} = \frac{\delta_{l_1 l_2} \delta_{m_1 m_2} (\tilde{C}_{l_1} - C_{l_1})}{\sqrt{C_{l_1} C_{l_2}}}, \quad (50)$$

$$\mathcal{B} = \frac{a_{l_1 m_1} a_{l_2 m_2}^* - C_{l_1 m_1, l_2 m_2}}{\sqrt{C_{l_1} C_{l_2}}}, \quad (51)$$

$$\mathcal{C} = \frac{C_{l_1 m_1, l'_1 m'_1} C_{l_2 m_2, l'_2 m'_2}}{\sqrt{C_{l_1} C_{l_2} C_{l'_1} C_{l'_2}}}, \quad (52)$$

where \tilde{C}_l is understood to be the canonical model including the predicted deviation. Also we now consider $C_{l_1 m_1, l_2 m_2} = \langle a_{l_1 m_1} a_{l_2 m_2}^* \rangle$. The corresponding isotropic basis functions are

$$\mathcal{R} = \frac{\delta_{l_1 l_2} \delta_{m_1 m_2}}{v_{l_1} v_{l_2}} R_{nl_1}, \quad (53)$$

which must be supplemented to incorporate anisotropic experimental effects such as cut-sky masking. We can exploit the tractability of the modal inverse covariance weighting in several ways. First we could consider constraining specific models with specific signatures in the power spectrum with a particular α_n , such as oscillatory models. Secondly we could use the method for full power spectrum estimation using the recovered β to iteratively improve the best fit theoretical model. This would entail linking to Monte Carlo simulations which explore cosmological parameter space. Finally, we need not restrict ourselves to isotropic models and instead using additional anisotropic basis functions, we can to constrain anisotropy in the CMB.

For large-scale structure and other 3D polyspectra estimation, this approach has an entirely analogous implementation discussed in ref. [21]. We simply replace our multipoles with the density perturbation so now we have

$$\langle \mathbf{a}_\emptyset \rangle \equiv \langle \delta_{\mathbf{k}_1} \delta_{\mathbf{k}_2} \dots \delta_{\mathbf{k}_n} \rangle \quad (54)$$

where now the vector $\boldsymbol{\varphi}$ represents $\boldsymbol{\varphi} = (\mathbf{k}_1, \mathbf{k}_2, \dots, \mathbf{k}_n)$ which cover the full $3n$ -dimensional domain of possible polyspectra. Here momentum conservation and isotropy imply we can make the reduction to the subspace $\varphi = (k_1, k_2, \dots, k_n)$ and the two are related by

$$\mathbf{a}_{\boldsymbol{\varphi}} = (2\pi)^3 \delta(\mathbf{k}_1 + \mathbf{k}_2 + \dots + \mathbf{k}_n) \mathbf{a}_{\varphi}, \quad (55)$$

The rest of the modal method then follows analogously with two main caveats. The first being that the growth of structure is non-linear so even with gaussian initial conditions we produce measurable polyspectra. The form of these can be calculated from N -body simulations and we can seek deviations from these due to primordial and other effects. Secondly there is a non-linear relationship between the late time and primordial polyspectra and so a Monte Carlo approach is required to find the parameter values which provide the best fit to the data. Error bars can then be obtained from simulations with the chosen primordial spectrum. However, the inverse covariance matrix can still be calculated from Gaussian simulations, so it does not need to be recalculated for differing levels of primordial non-Gaussianity.

V. CONCLUSION

We have shown that full inverse covariance weighting can be naturally incorporated in the modal polyspectra estimator to achieve optimality. The resulting modal covariance matrix can be efficiently calculated and inverted. This greatly reduces the computational cost and memory requirements of polyspectra estimation, making optimal analysis of high resolution experiments like Planck possible. We emphasise that the primary efficiencies discussed here arise because we have effectively projected out all but the modal degrees of freedom needed to characterise the model (or models) under investigation, while retaining an adequate description of the noise and systematics with which it is correlated.

We have only briefly discussed the many applications of this work. These include improved WMAP7 constraints on a wide variety of theoretical models, together with a model-independent constraint on the overall magnitude of the bispectrum in ref. [3]. A principal component analysis can be used to identify the noise, mask and other shapes for preconditioning the inverse covariance weighting. A forthcoming paper will use eigenmodes to characterise and marginalise over foregrounds and other contaminants in preparation for the CMB analysis of Planck data. The analysis of higher order correlators in three dimensions represents a particular challenge as many billion galaxy surveys become available. The 3D modal estimator [21] allows for efficient polyspectra estimation and for the setting of arbitrary non-Gaussian initial conditions. Combined with inverse covariance weighting, this methodology may help unravel the many systematic and evolutionary effects present in large-scale structure observations.

VI. ACKNOWLEDGEMENTS

We are very grateful for many useful conversations with Michele Liguori and Donough Regan with whom we have developed the modal bispectrum methodology described here. The inverse covariance weighting aspect of the work has been taken forward with Helge Gruentjen and we are very grateful for numerous enlightening discussions. We also thank Anthony Challinor for useful comments on an earlier draft of this paper. We are also grateful for helpful conversations and interactions with Eugene Lim, Hiro Funakoshi, Marcel Schmittfull and Xingang Chen.

Simulations were performed on the COSMOS@DiRAC supercomputer (an SGI Altix UV1), and the national DiRAC facility is funded by STFC and DBIS. JRF and EPS were supported by STFC grant ST/F002998/1 and the Centre for Theoretical Cosmology.

Appendix A: Anisotropic noise and bispectrum optimality

Anisotropic instrument noise provides an explicit example which illustrates some of the points raised above, notably the nature of its contribution to the variance the importance of the linear term in (5) for

optimal estimation. WMAP non-Gaussian analysis to date has used a simple instrument noise model which assumes no correlations between pixels. If n_{lm} is the harmonic transform of the pixel-variance map, then the covariance of the noise contribution to the CMB temperature is given by

$$\langle a_{l_1 m_1}^N a_{l_2 m_2}^N \rangle = \frac{4\pi}{n_{\text{pix}}} \sum_{lm} (-1)^m \mathcal{G}_{m_1 m_2 - m}^{l_1 l_2 l} n_{lm}. \quad (\text{A1})$$

(See, for example, the discussion of covariance under rotations in [22].) The covariance of the cut-sky mask also can be represented in this form by taking the masked pixels to have a very large noise so that the covariance weighting neglects that section of the map. We shall now consider the contributions to the variance of the estimator both with and without the linear term.

We assume CMB signal plus instrument noise can be represented as $a_{lm} + a_{lm}^N$ with respective covariances $C_{l_1 m_1, l_2 m_2}^S = \langle a_{l_1 m_1} a_{l_2 m_2} \rangle = (-1)^{m_1} C_{l_1} \delta_{l_1 l_2} \delta_{m_1 - m_2}$ and $C_{l_1 m_1, l_2 m_2}^N = \langle a_{l_1 m_1}^N a_{l_2 m_2}^N \rangle$. Again restricting attention to the isotropic subspace by expanding using isotropic modes (27), the modal covariance matrix with the inhomogeneous noise takes the form

$$\begin{aligned} \zeta_{nn'} &= \frac{1}{6} \sum_{l_i m_i l'_i m'_i} R_{n l_1 l_2 l_3} \mathcal{G}_{m_1 m_2 m_3}^{l_1 l_2 l_3} C_{l_1 m_1, l'_1 m'_1} C_{l_2 m_2, l'_2 m'_2} C_{l_3 m_3, l'_3 m'_3} \mathcal{G}_{m'_1 m'_2 m'_3}^{l'_1 l'_2 l'_3} R_{n' l'_1 l'_2 l'_3} \\ &= \frac{1}{6} \sum_{l_i m_i l'_i m'_i} R_{n l_1 l_2 l_3} \mathcal{G}_{m_1 m_2 m_3}^{l_1 l_2 l_3} (C^S C^S C^S + 3C^S C^S C^N + 3C^S C^N C^N + C^N C^N C^N) \mathcal{G}_{m'_1 m'_2 m'_3}^{l'_1 l'_2 l'_3} R_{n' l'_1 l'_2 l'_3} \\ &= \frac{1}{6} \sum_{l_i l'_i} R_{n l_1 l_2 l_3} R_{n' l'_1 l'_2 l'_3} \left[h_{l_1 l_2 l_3}^2 \delta_{l_1, l'_1} \delta_{l_2, l'_2} \delta_{l_3, l'_3} \left(C_{l_1} C_{l_2} C_{l_3} + C_{l_1} C_{l_2} \frac{n_{00}}{\sqrt{4\pi}} \right) \right. \\ &\quad + \sum_l h_{l_1 l_2 l_3} h_{l_1 l'_2 l'_3} h_{l_2 l'_2 l'_3} h_{l_3 l'_3 l'_3} \left\{ \begin{matrix} l_1 & l_2 & l_3 \\ l & l'_3 & l'_2 \end{matrix} \right\} C_{l_1} C_l^N \\ &\quad \left. + \sum_{l' l''} h_{l_1 l_2 l_3} h_{l'_1 l'_2 l'_3} h_{l_1 l'_1 l'_1} h_{l_2 l'_2 l'_2} h_{l_3 l'_3 l''} \left\{ \begin{matrix} l_1 & l_2 & l_3 \\ l'_1 & l'_2 & l'_3 \\ l & l' & l'' \end{matrix} \right\} b_{l' l''}^N \right] \end{aligned} \quad (\text{A2})$$

where C_l^N represents the isotropic power from the pixel-variance map,

$$C_l^N \equiv \sum_m \frac{(-1)^m}{2l+1} n_{lm} n_{l-m}, \quad (\text{A3})$$

$b_{l_1 l_2 l_3}^N$ is the isotropic part of its bispectrum,

$$b_{l_1 l_2 l_3}^N = \sum_{m_i} \mathcal{G}_{m_1 m_2 m_3}^{l_1 l_2 l_3} n_{l_1 m_1} n_{l_2 m_2} n_{l_3 m_3}. \quad (\text{A4})$$

Standard identities were used to convert summed products of four and five Wigner-3j symbols into the Racah-6j and Racah-9j symbols respectively.

The expression (A2), which yields that the isotropic covariance from anisotropic noise (A1), depends solely on the two- and three-point correlators of the pixel-variance map (i.e. the isotropic functions C_l^N and $b_{l_1 l_2 l_3}^N$). The reduction to isotropy means that the product form of the full covariance matrix \mathcal{C} can be reduced from a $l_{\text{max}}^6 \times l_{\text{max}}^6$ matrix to $l_{\text{max}}^3 \times l_{\text{max}}^3$ but this is still intractable. Here, however, we have a modal covariance matrix of only $n_{\text{max}} \times n_{\text{max}}$ which can be easily inverted.

Now consider the contribution from inhomogeneous noise to the covariance matrix in the absence of a linear term in the estimator (31) but without the linear term. In this case, the full covariance has the

following additional cross-terms

$$\begin{aligned}
\langle \beta_n \beta_{n'} \rangle_\times &\propto 9 \sum_{l_i m_i l'_i m'_i} R_{n l_1 l_2 l_3} \mathcal{G}_{m_1 m_2 m_3}^{l_1 l_2 l_3} C_{l_1 m_1, l_2 m_2} C_{l_3 m_3, l'_3 m'_3} C_{l'_1 m'_1, l'_2 m'_2} \mathcal{G}_{m'_1 m'_2 m'_3}^{l'_1 l'_2 l'_3} R_{n' l'_1 l'_2 l'_3} \\
&= \sum_{l_i m_i l'_i m'_i} R_{n l_1 l_2 l_3} \mathcal{G}_{m_1 m_2 m_3}^{l_1 l_2 l_3} [9C^S C^S C^S + 3(6C^S C^S C^N + 3C^S C^N C^S) \\
&\quad + 3(3C^N C^S C^N + 6C^S C^N C^N) + 9C^N C^N C^N] \mathcal{G}_{m'_1 m'_2 m'_3}^{l'_1 l'_2 l'_3} R_{n' l'_1 l'_2 l'_3} \\
&= 9 \sum_{l_i l'_i} R_{n l_1 l_2 l_3} R_{n' l'_1 l'_2 l'_3} \left[\frac{(2l_1 + 1)(2l'_1 + 1)}{4\pi} \delta_{l_1, l_2} \delta_{l'_1, l'_2} \delta_{0, l_3} \delta_{0, l'_3} \left(C_{l_1} C_{l'_1} C_0 + 2C_{l_1} C_0 \frac{n_{00}}{\sqrt{4\pi}} + C_{l_1} C_{l'_1} \frac{n_{00}}{\sqrt{4\pi}} \right) \right. \\
&\quad + \frac{(2l_1 + 1) h_{l_1 l_2 l_3}^2 h_{l'_1 l'_2 l'_3}^2}{4\pi} (-1)^{l_1} \delta_{l_1, l_2} \delta_{0, l_3} C_{l_1} C_{l'_3}^N + \frac{2h_{l_1 l_2 l_3}^2 h_{l'_1 l'_2 l'_3}^2}{(2l_3 + 1)} \delta_{l_3, l'_3} C_{l_3} C_{l'_3}^N \\
&\quad \left. + \sum_l \frac{h_{l_1 l_2 l_3}^2 h_{l'_1 l'_2 l'_3}^2}{(2l_3 + 1)(2l'_3 + 1)} b_{l_3 l'_3 l} \right] \tag{A5}
\end{aligned}$$

The first line of the resulting expression corresponds to a monopole so can be ignored. The remaining two lines containing three cross terms do not in general vanish, but once again they contain only contributions from the isotropic parts of the pixel-variance map correlators C_l^N and $b_{l_1 l_2 l_3}^N$. Other effects which break rotational invariance in a similar manner, such as a cut-sky mask, can be treated similarly.

It is important to note that the existence of the non-zero cross-terms (A5) means that the minimum variance cannot be achieved without use of the linear term as shown in (33).

-
- [1] J. R. Fergusson and E. P. S. Shellard, Phys. Rev. **D76**, 083523 (2007), astro-ph/0612713.
 - [2] J. R. Fergusson, M. Liguori, and E. P. S. Shellard (2009), 0912.5516.
 - [3] J. R. Fergusson, M. Liguori, and E. P. S. Shellard (2010), 1006.1642.
 - [4] J. R. Fergusson, D. M. Regan, and E. P. S. Shellard (2010), 1012.6039.
 - [5] K. M. Gorski, Astrophys.J. **430**, L85 (1994), fermilab Library Only, astro-ph/9403066.
 - [6] J. Bond, Phys.Rev.Lett. **74**, 4369 (1995), astro-ph/9407044.
 - [7] S. P. Oh, D. N. Spergel, and G. Hinshaw, Astrophys.J. **510**, 551 (1999), astro-ph/9805339.
 - [8] A. F. Heavens, MNRAS **299**, 805 (1998), arXiv:astro-ph/9804222.
 - [9] K. M. Smith, O. Zahn, and O. Dore, Phys.Rev. **D76**, 043510 (2007), 0705.3980.
 - [10] K. M. Smith, L. Senatore, and M. Zaldarriaga, JCAP **0909**, 006 (2009), 0901.2572.
 - [11] F. Elsner and B. D. Wandelt, Astron.Astrophys. **540**, L6 (2012), 1202.4898.
 - [12] J. R. Fergusson and E. P. S. Shellard, Phys. Rev. **D80**, 043510 (2009), 0812.3413.
 - [13] A. Curto, E. Martinez-Gonzalez, and R. Barreiro (2010), * Temporary entry *, 1007.2181.
 - [14] M. Bucher, B. Van Tent, and C. S. Carvalho (2009), 0911.1642.
 - [15] O. Rudjord, F. K. Hansen, X. Lan, M. Liguori, D. Marinucci, et al., Astrophys.J. **701**, 369 (2009), * Brief entry *, 0901.3154.
 - [16] P. Creminelli, A. Nicolis, L. Senatore, M. Tegmark, and M. Zaldarriaga, JCAP **0605**, 004 (2006), astro-ph/0509029.
 - [17] A. Curto, E. Martinez-Gonzalez, and R. B. Barreiro, ArXiv e-prints (2011), 1111.3390.
 - [18] D. Babich, Phys. Rev. **D72**, 043003 (2005), astro-ph/0503375.
 - [19] E. Komatsu et al. (WMAP) (2008), 0803.0547.
 - [20] D. M. Regan, E. P. S. Shellard, and J. R. Fergusson (2010), 1004.2915.
 - [21] J. R. Fergusson, D. M. Regan, and E. P. S. Shellard (2010), 1008.1730.
 - [22] D. Hanson, A. Lewis, and A. Challinor, Phys. Rev. **D81**, 103003 (2010), 1003.0198.
 - [23] E. Komatsu, D. N. Spergel, and B. D. Wandelt, Astrophys. J. **634**, 14 (2005), astro-ph/0305189.
 - [24] S. Donzelli, F. K. Hansen, M. Liguori, D. Marinucci, and S. Matarrese, ArXiv e-prints (2012), 1202.1478.